

METADATA AGGREGATION:

OAI-PMH: IMLS Digital Collections & Content (2002-07)

Library-GSLIS research indexing digital collections



Primary research areas:

- Early testing & assessment of the Open Archives Initiative Protocol for Metadata Harvesting
- Refining approaches to collection-level description
- Importance & usefulness of collection-level description
- Trends & patterns in metadata creation practice
- Best practices for shareable metadata & OAI-PMH implementation
- Diffusion of best practices into IMLS community
- Scope:
 - 150+ collections
 - 250,000+ item-level record
 - funding from IMLS NLG & LSTA s

METASEARCH DEVELOPMENT

Broadcast search and retrieval spanning multiple databases, sources, platforms, protocols, and vendors all at once.

- Libraries typically spend over half of their materials budgets on electronic resources licensed from publishers, vendors, information centers.
- Users face a bewildering array of distributed, discrete, heterogeneous online information resources.
- Many of these resources are not available through Web search engines.
- Libraries are using metasearch techniques to simultaneously search and access these discrete information resources.

Sustainable Access to Research & Scholarship

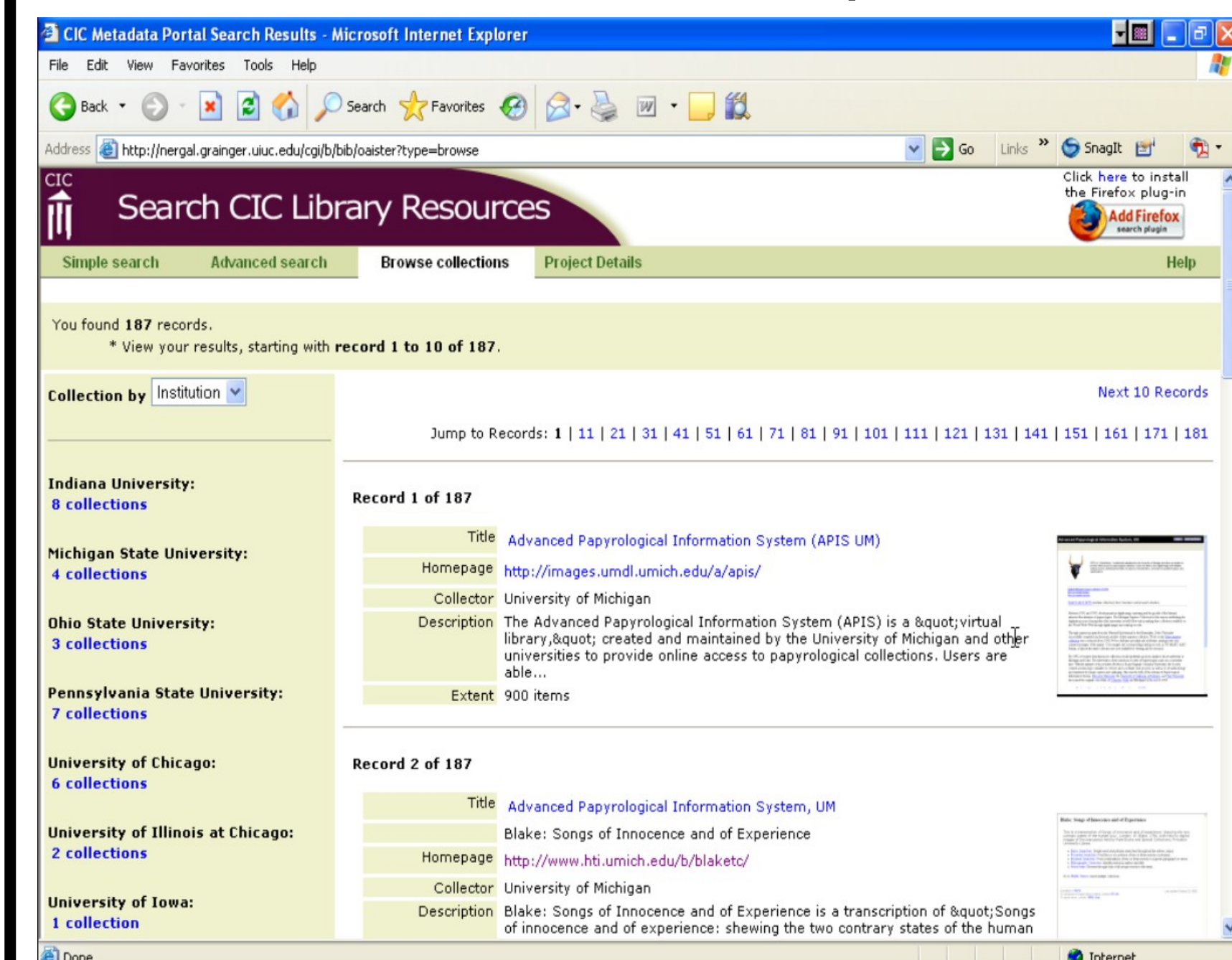
- **Systematic dissemination** of research, data, and scholarship in all formats
- **Open or restricted access** to material
- **Persistent URLs** (using Handle technology)
- **Full text searching** of most documents
- **Long term preservation** of deposited research
- **Support services** for large scale deposits, format preservation, and author right consultations
- **On Track for TRAC** (Trustworthy Repository Audit Checklist-NARA)
- **Contact:** Sarah Shreeves, IDEALS Coordinator

sshreeve@uiuc.edu



OAI-PMH: CIC Metadata Portal (2003-06)

A project of the research libraries of the Committee on Institutional Cooperation



Primary research areas:

- New models for CIC Library collaboration & interoperability
- Interface design & testing across multiple CIC sites.
- Analyses of how metadata design impacts on real end-user searching
- Experimentation with mixed levels of item description (MARC, MODS, QDC, DC)
- Scope:
 - 200+ collections
 - 500,000+ item-level records

Custom Metasearch Engine Features

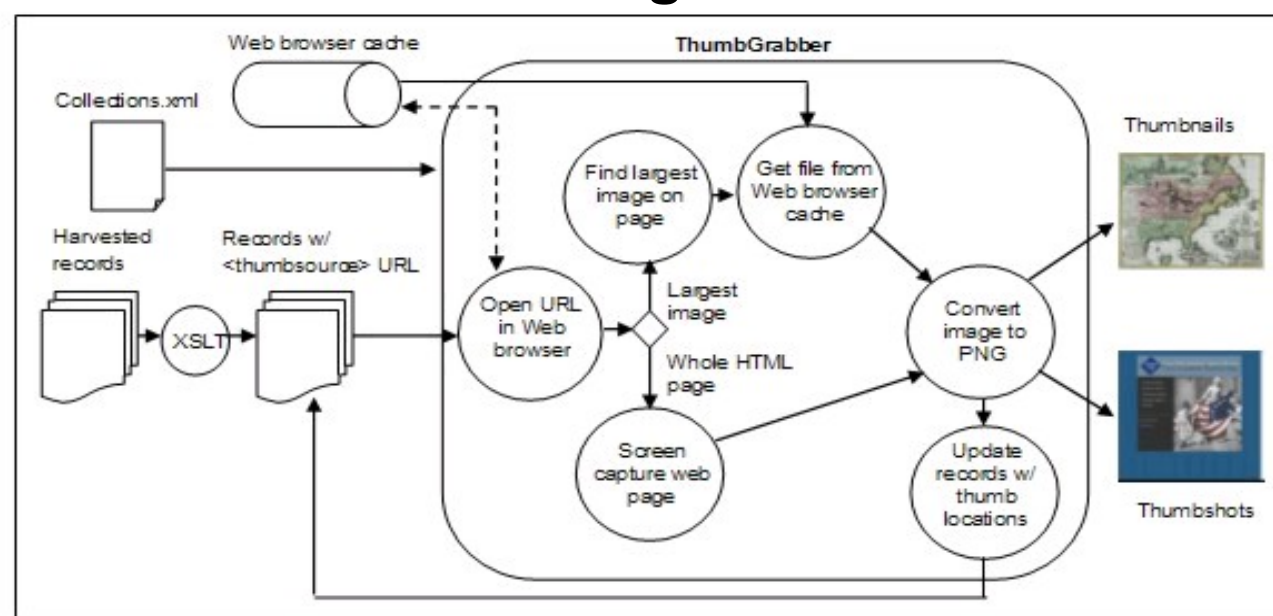
- "Spell checking of user search arguments with clickable spelling corrections;
- "Search term stopword removal & table-driven query term switching;
- "Guided search assistance and navigational prompts;
- "Specified target result display removal when there are no hits;
- "Software controlled database selection algorithms;
- "Links into the native interface at the point of completed search for further user searching.

Repository Infrastructure

- DSpace open-source software (MIT & HP)
 - Java, PostgreSQL database, Tomcat web server
 - Wide acceptance (200+ institutions, 45+ countries)
 - Persistent URLs (using CNRI Handle System)
- Customizations/Localizations
 - OpenOffice.org format conversions (e.g. Word to PDF)
 - Bluestem NetID login
 - UIUC Active Directory Groups
- Server Infrastructure
 - Managed by CITES System Management Group
 - Expandable 6TB SAN
 - Off-campus backups

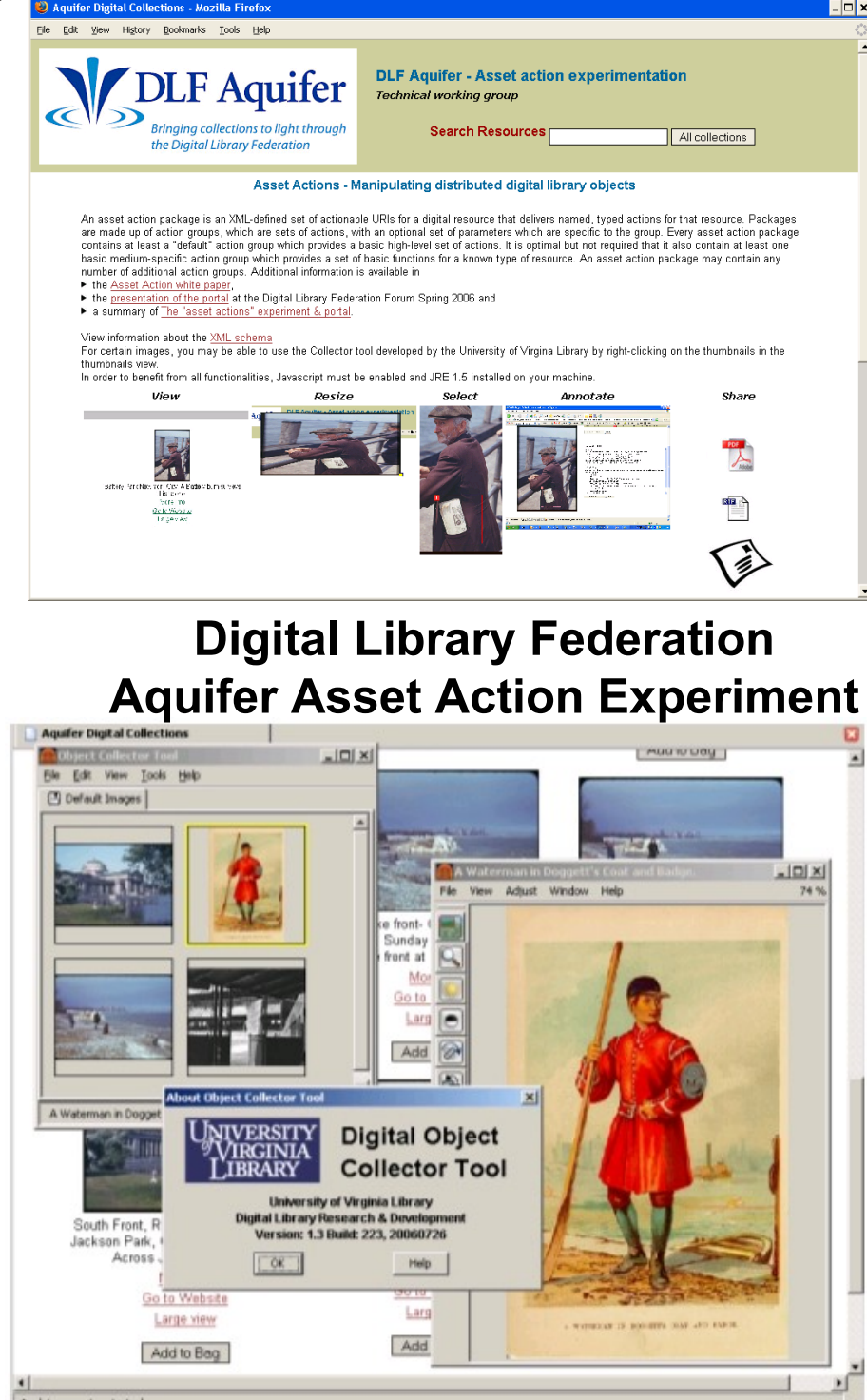
Other Applications of OAI-PMH

Automated Capture of Thumbnail Images



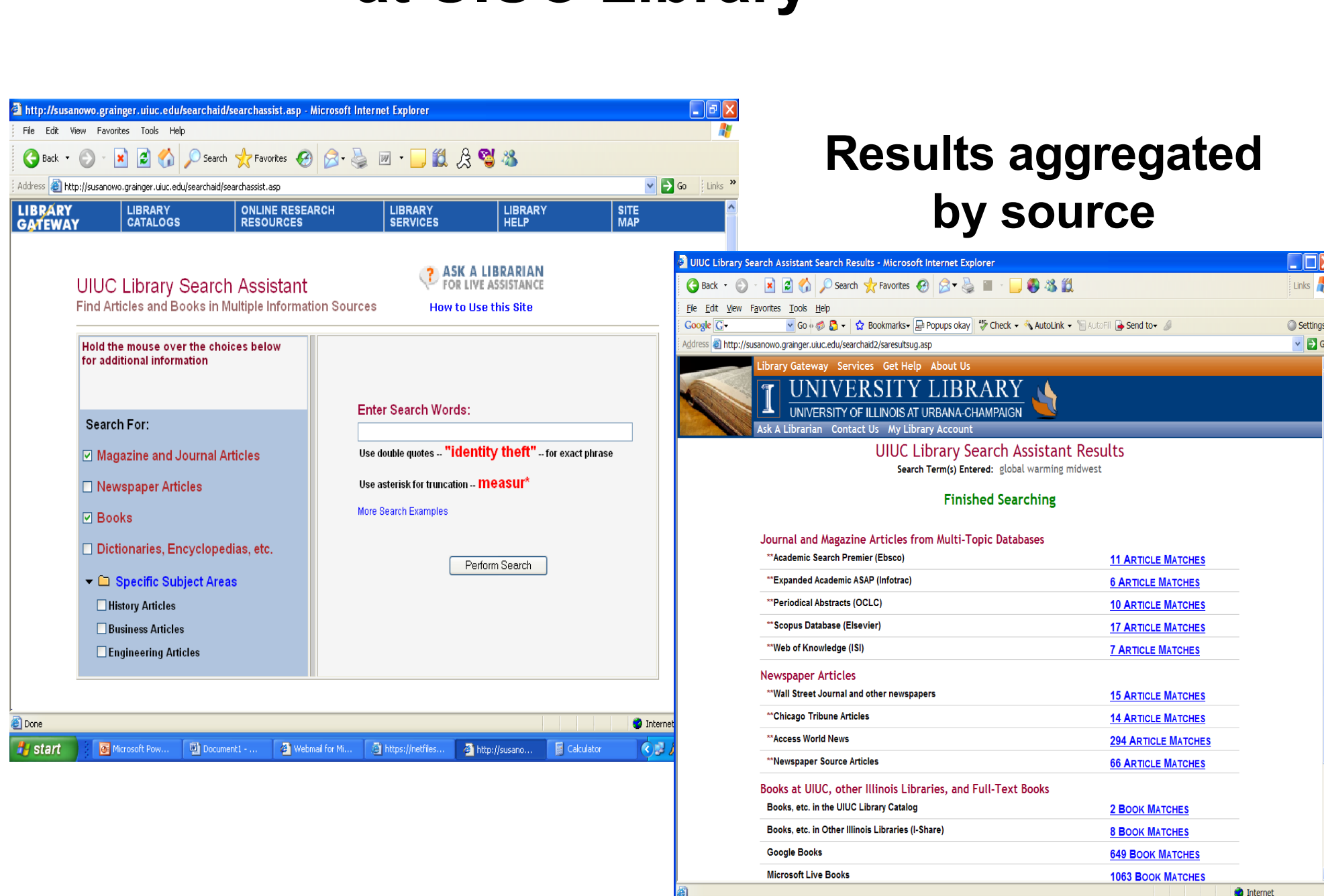
Using collection-level description to enhance item retrieval

Partial match	# of queries with at least 1 item-level match of the case	% of queries with at least 1 item-level match of the case
Case A	287	17.00%
Case B	21	1.20%
Case C	761	45.10%
Case D	25	1.50%
Case E	20	1.20%
Case F	222	13.10%
Case G	1,639	97.00%
Case H	940	55.70%
Case J	945	56.00%



Applications of Metasearch Software at UIUC Library

Search Assistant: searches multiple hidden Web resources

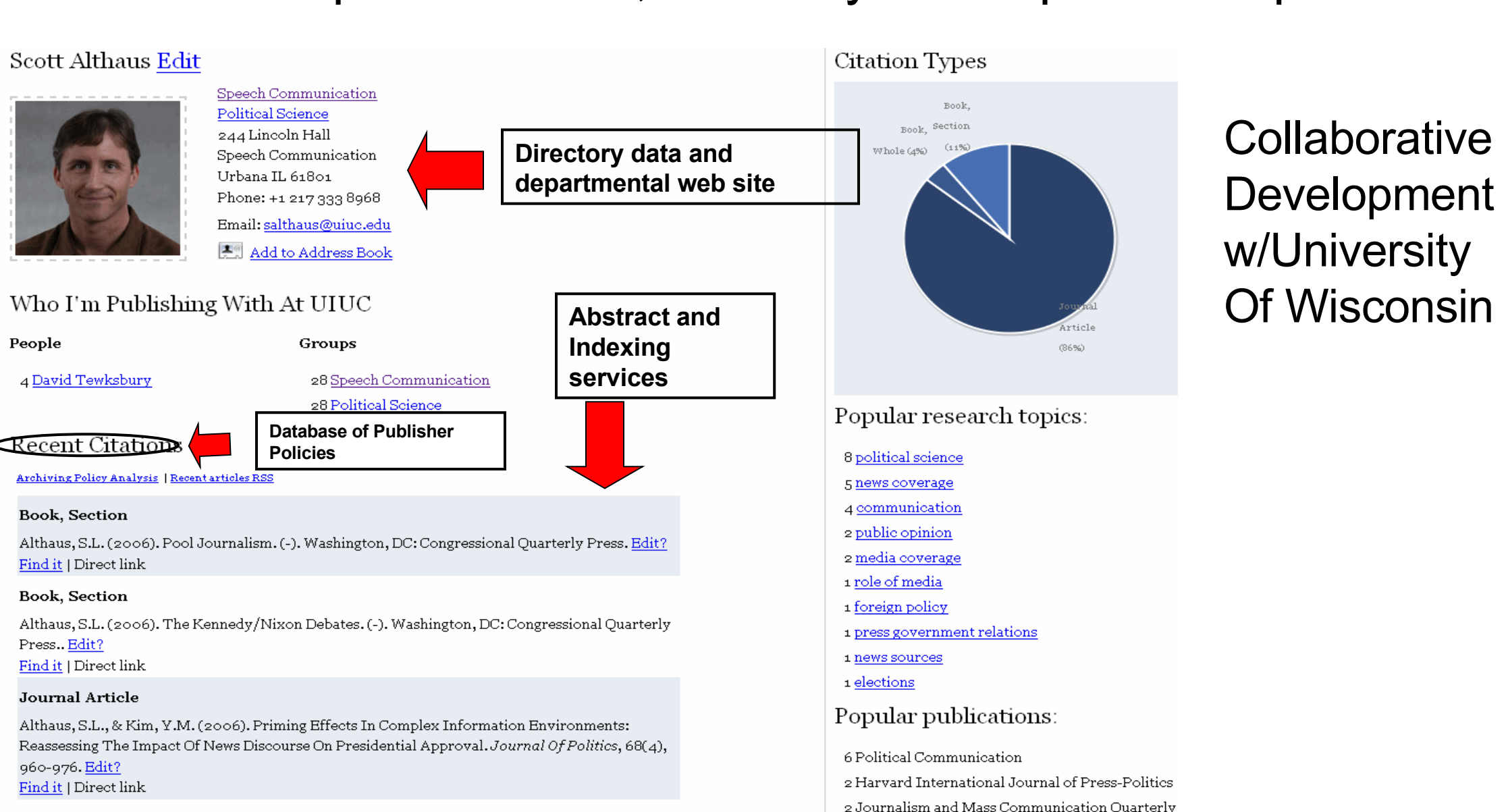


Funded by NSF, Mellon

Services Layered Over Repositories:

BibApp: Understanding Faculty Publishing

A mash-up of citations, directory info & publisher policies



Collaborative Development w/University Of Wisconsin

